

## CONVERSIÓN DE LOS REGISTROS AUTOMATIZADOS DE BIBLIOGRAFÍA ESPAÑOLA Y SU INTEGRACIÓN EN LA BASE DE DATOS ARIADNA\*

PILAR DOMÍNGUEZ SÁNCHEZ  
*Biblioteca Nacional (Madrid)*

### INTRODUCCIÓN

Los primeros intentos de automatización de Bibliografía Española se llevaron a cabo en 1969, aunque fue en el año 1971 cuando se elaboró un formato para la preparación de cintas de fotocomposición con el fin de publicar el boletín en papel de la bibliografía nacional de monografías, cuyo primer número elaborado con este sistema salió en enero de 1972. Esta labor la realizó el Instituto Bibliográfico Hispánico hasta el año 1985, en que, desaparecido el citado organismo, quedaron integradas sus funciones en la Biblioteca Nacional.

Desde 1976 se utilizó el formato basado en la edición IBERMARC de dicho año, que era el resultado de una adaptación del MARC empleado por la Library of Congress. Esta primera edición tuvo dos revisiones posteriores, en el año 1981 y en 1988, cuyos cambios se vieron reflejados en la estructura de la información contenida en las cintas.

Estos primeros registros automatizados, que fueron preparados por el Instituto Bibliográfico Hispánico, suman algo más de 300.000 y comprenden las fechas de 1976 a 1987.

En 1987 entró en funcionamiento el primer sistema automatizado de la Biblioteca Nacional, denominado SABINA. En este sistema la catalogación se realizaba en línea y la edición de la Bibliografía Española pasó a ser un subproducto del sistema. Es en 1991 cuando se produjo la migración a un nuevo sistema, actualmente en uso, denominado ARIADNA, que cuenta con nuevas funcionalidades y aplicaciones, entre ellas, un programa de integración de registros.

---

\* Comunicación leída en el 59 Consejo y Conferencia General de la IFLA, Barcelona, agosto 1993.

Paralelamente a este proceso, en 1989 se iniciaron los proyectos de reconversión del catálogo manual de la Biblioteca Nacional. Y, una parte de esta reconversión se apoyó en la utilización de las cintas de fotocomposición existentes.

Los registros creados durante los años 1976 a 1987 estaban prácticamente sin documentar. Aparentemente estos registros estaban en formato MARC, pero la realidad era bien distinta. En general, la situación de los registros era muy variada, debido entre otras cosas a que las codificaciones IBERMARC de los registros no eran homogéneas. A lo largo de estos años, el formato, las puntuaciones y el juego de caracteres se iba adaptando a las necesidades y servidumbres que surgían al preparar el contenido de las cintas para su edición en papel. Es importante insistir que esta información no estaba elaborada para formar parte de una base de datos.

Atendiendo a estas circunstancias y dada la indudable utilidad de estos registros bibliográficos, tanto para una labor de cooperación como para un mayor rigor en el control bibliográfico nacional, era fundamental acometer esta tarea de conversión.

Por otra parte y teniendo en cuenta el crecimiento cada vez mayor de la producción bibliográfica y la necesidad de satisfacer las constantes demandas de los usuarios, no se podía prescindir de esta información para dar a conocer nuestra bibliografía, facilitar la información bibliográfica y cubrir el más amplio campo de necesidades dentro del control bibliográfico nacional.

La Biblioteca Nacional de España es en estos momentos la responsable de elaborar la bibliografía nacional y los registros sobre los que se ha trabajado en este centro para su conversión son los realizados a partir de 1976.

#### FASES DE PROCEDIMIENTO

Teniendo en cuenta que los registros abarcaban un período de 11 años, en los que el panorama de la normativa bibliográfica ha sufrido algunos cambios, el primer paso era el análisis del estado de la información, que incluía el estudio de la codificación y estructura MARC, así como del juego de caracteres.

El procedimiento utilizado fue el de obtener en papel esos registros en hexadecimal, utilizando un programa de IBM denominado DITO.

Este programa permitía establecer:

- Código hexadecimal utilizado en los registros;
- Delimitadores de los códigos de subcampos, y
- Códigos utilizados para fin de campo y para fin de registro.

Después pasaban por un filtro muy estricto, y aquí se comprobaban tres aspectos fundamentales:

- Campos y subcampos existentes;
- Obligatoriedades, y
- Repetibilidades.

En cuanto a los indicadores no se comprobaba su contenido, pero sí la estructura.

Con todo ello se obtenían estadísticas y listas para conocer el estado de los registros, campos, etc. y poder posteriormente tratar la información.

Como resultado del análisis, se comprobó que un 50% de los registros cumplían los requisitos para ser integrados en la base de datos.

Dentro de ARIADNA, en el subsistema de integración, permanecía la información que no había pasado a la base de datos, es decir, mantenía aquellos registros que no habían cumplido los requisitos del filtro, para de este modo trabajar con esos registros rechazados.

Este módulo permite el tratamiento masivo de la información y es útil, por tanto, para la depuración antes de entrar en la base de datos. En esta fase se revisa el tratamiento que debe darse a la información, mediante la comprobación de los campos que faltan y su obligatoriedad; el cambio masivo del contenido de los indicadores; la corrección de puntuaciones (y/o ortografía), y la agregación o supresión de valores.

A partir de esta revisión, se decide lo que se puede corregir por programa y lo que tiene que ser modificado manualmente.

#### TRADUCCIÓN DEL FORMATO

Con la información obtenida en las fases que acabo de mencionar se planteó el tema de la conversión de formatos.

El mapping era necesario realizarlo a dos niveles:

Primero hay que traducir ese formato MARC al IBERMARC, así como el juego de caracteres al nuestro de la Biblioteca Nacional, y en segundo lugar, traducir el formato IBERMARC a ARIADNA para poder trabajar posteriormente con esa información en la base de datos.

Para una parte de los registros, los procedentes de SABINA, el proceso de transformación de la información se vio facilitado, ya que desde el primer momento se habían realizado siguiendo las cinco normas ISO: la 2709 desde el punto de vista de la estructura lógica; la 1001 desde el punto de vista de la estructura física y las tres de juego de caracteres: la 646, la 5426 y la 6630.

Es fundamental señalar que todas las transformaciones pueden conlle-

var pérdida de información o modificación inexacta de registros, cuando se ha avanzado en una estructura nueva. Lo importante no estriba en la dificultad del trabajo, sino en tener perfectamente controlado cuál es el estado de la cuestión de los registros en cada una de las fases. Para ello, dentro del formato MARC existen unos campos, como son el 005, donde se indica la fecha de la última modificación, y también las seis primeras posiciones del 008, donde aparece señalada cuál es la entrada del registro en el fichero. De este modo, podemos saber si el registro ha sido modificado, cuándo y el motivo de la modificación.

#### DESTINO DE LA INFORMACIÓN

Una vez comentado el origen de la información, también es conveniente señalar el estado de la cuestión en cuanto se refiere al destino de esa información.

La Biblioteca Nacional, además de elegir una nueva solución informática, ha llevado a cabo la tarea de definir su propio formato y ha realizado la quinta edición del formato IBERMARC de monografías, así como la primera edición del formato IBERMARC de autoridades, lo que supone la posibilidad de tener por primera vez un control sistemático de los puntos de acceso de forma unificada.

#### *Características de la base de datos ARIADNA: estructura de la información*

Esta base de datos permite al usuario definir la estructura de datos, la cual está diseñada siguiendo el modelo Entidad/Relación, dando lugar al procedimiento denominado navegación.

En ARIADNA existen tres clases de registros: los de autoridad, los bibliográficos y los de fondos y localizaciones.

Dentro de cada uno de estos grupos, en ARIADNA se han creado distintos formatos que reciben el nombre de «entes», entendiéndolo como un ente aquel objeto del mundo real del que se quiere guardar información y que posee unas características concretas que lo identifican y que, a su vez, lo diferencian de los demás. La información que constituye un ente se descompone en elementos o unidades mínimas de información y, estos elementos se agrupan entre sí formando grupos homogéneos. Estos grupos son análogos a los campos MARC.

Por tanto cada ente posee los elementos necesarios para crear un determinado tipo de registro. En ARIADNA se han definido, por ejemplo, tan-

tos entes de autoridad como tipos de autoridades: nombre de persona, de entidad, de congreso, etc.

En cuanto a los registros bibliográficos, existen tantos entes bibliográficos como materiales documentales: entes de monografías modernas, monografías antiguas, manuscritos, publicaciones seriadas, material cartográfico, etc.

La definición se ha hecho dentro del marco de:

- Reglas de catalogación
- ISBD y
- La estructura del formato IBERMARC (que es muy similar al US-MARC).

Otro aspecto importante, además de los objetos o entidades de la base de datos, es el de las relaciones entre esos entes.

Las relaciones que se pueden establecer en los registros se denominan asociaciones y éstas pueden ser:

- Implícitas, cuando parte de la información de un registro, el localizador, pasa a formar parte de otro, y
- Explícitas, cuando el catalogador da de alta un registro de asociación, que consta de dos elementos como mínimo, existiendo un localizador por cada uno de los dos entes entre los que se quiere establecer la asociación.

El modelo de datos, por tanto, hace referencia, por un lado al conjunto de entidades con sus atributos y, por otro, al conjunto de asociaciones entre ellas.

Con esta estructura en la base de datos fue necesario proceder a la fragmentación de la información a tres niveles, según a qué tipo de registro fuera destinada: registros bibliográficos, de autoridades y de fondos y localizaciones.

Como la información en las cintas de bibliografía española no se encontraba así, ya que existía un registro bibliográfico por cada unidad bibliográfica (item), ha sido necesario realizar un sangrado de las autoridades a partir de los campos 1XX, 6XX y 7XX de cada registro bibliográfico, para crear un registro de autoridad aceptable en la base de datos (ARIADNA).

El procedimiento seguido es moderno y consiste en la extracción de los distintos puntos de acceso identificados como tal en el registro bibliográfico y la transformación en registros independientes de autoridades, es decir, un campo 100 de un registro bibliográfico —un encabezamiento de nombre de persona— se convierte en un registro de autoridad mediante el

sangrado de ese campo para formar un registro nuevo, colocando un puntero en el registro de origen y añadiendo una serie de valores por defecto, como son un leader, un 001, un 005, un 008 y un 040. Posteriormente esos registros se validan contra sí mismos.

#### PROBLEMÁTICA DE LA CONVERSIÓN

En la conversión de los registros automatizados surgieron diversos problemas.

Los referidos a las autoridades serían los siguientes:

- Unión del registro bibliográfico con el registro de autoridad
- Alternativos (asociación)
- Codificación correcta
- Autoridades bien sangradas
- Autoridades bien enganchadas entre sí
- Validación de los registros de autoridad contra sí mismos, existiendo muchos de no validación, aunque parezca imposible.

Como herencia de la ficha manual, desde el punto de vista de lo que es una base de datos, podemos decir que es problemático el hecho de tener que diferenciar un autor principal de un secundario. La estructura manual hace que esto lo tratemos de forma distinta. El formato MARC americano presenta por ejemplo indicadores distintos para un campo 700 que para 100, lo que hace que, cuando se cruzan las informaciones, se produzcan importantes disparidades. Mayores problemas se presentan, si la ficha es de autor-título, ya que, por ejemplo, cuando incluimos un título en el 700 como título prácticamente uniforme, no le ponemos indicadores a ese título y, por tanto, cuando confrontamos un mismo autor con el mismo título en una entrada principal y en una secundaria se produce toda una serie de modificaciones.

De todo esto se deduce que para el control de autoridades es necesario asegurar que los nombres, títulos uniformes y encabezamientos de materia que se utilizan en los registros sean únicos, uniformes y correctamente formulados.

La única garantía reside en revisar uno a uno cada encabezamiento, ya que en ocasiones encontramos los nombres de personas que aparecen con el mismo nombre tratándose de autores diferentes y por otro lado hay variantes de un nombre asociado a un autor.

En el primer caso, será necesario crear un modelo con unas características para distinguir autores, atendiendo a la fecha en que comenzó a publicar y la fecha de nacimiento, la lengua en que publica, el tema de la publicación y los títulos de las obras que ha publicado.

Por lo que se refiere a los encabezamientos de materia, es preciso corregir los errores en los códigos de subcampo, así como igualmente corregir la utilización incorrecta de las etiquetas. También nos encontramos con la existencia de encabezamientos de materia con el mismo contenido pero en orden diferente, así como encabezamientos obsoletos. En este último caso las correcciones se hacen a nivel del encabezamiento.

El sistema de la Biblioteca Nacional facilita una buena recuperación permitiendo la posibilidad ya mencionada de navegación, lo cual hace más flexible el fichero de autoridades, y facilita el control y la coherencia de las mismas.

Otro problema importante ha sido el juego de caracteres, pues los registros convertidos tenían juegos de caracteres dispares y con falta de coherencia. Además, la mayoría no se atenía a los estándares. Su procedencia era ASCII, pero un ASCII defectuoso. Fue necesario realizar un estudio detallado de la información, ya que algunas combinaciones de bits tenían significados polivalentes y, posteriormente, crear un programa a la medida para cada información.

La Biblioteca Nacional ha decidido un juego de caracteres propio, que está basado en el EBCDIC. Lo cual no es obstáculo para que la información se exporte de una manera completamente normalizada.

Por lo que se refiere a los registros bibliográficos, hay que decir que en ARIADNA el registro bibliográfico contiene toda la información asociada con una edición específica de un trabajo. Cada uno de ellos constituye una entidad (ente) diferente formado por una serie de elementos (campos y subcampos). Pero aquí también nos encontramos con una cantidad considerable de duplicados a tenor de la estructura de la base de datos, ya que los registros convertidos, cuando se crearon, no se asociaban a una edición, sino a la unidad física que se catalogaba y las obras en volúmenes, tal y como iban llegando. Por tanto, en muchos casos existía una obra duplicada con los volúmenes sueltos y otras veces completa. Ahora, toda la información del ejemplar se integra en el formato de fondos y localizaciones.

Además, por lo que se refiere a los duplicados, existen desde errores tipográficos hasta codificaciones diferentes de los campos, así como inconsistencias debidas a incoherencias tanto de juicios como de normas catalográficas. Para intentar corregir estas incoherencias y errores se está creando una serie de programas que permitan chequear en masa registros que se agruparán por tener ciertas similitudes.

Para aquellos errores que el programa puede reconocer, pero no corregir, será necesaria la revisión manual.

## CONCLUSIONES

La primera recomendación que surge como consecuencia para llevar a cabo un trabajo práctico es procurar que la información del formato esté plenamente documentada y que sepamos a qué formato nos referimos, con una mención también muy clara del juego de caracteres que se utiliza, para evitar muchos problemas posteriores como los que nos hemos encontrado, siendo necesario, por último, diferenciar muy claramente la estructura física de los registros de la estructura lógica.

No hay que olvidar que los procesos de conversión no hacen que mejore la calidad de la información, pero sí permite que la información existente sea coherente entre sí.

Por último manifestar que la conversión de estos registros será de gran utilidad para apoyar la reconversión de las bibliotecas españolas.

## BIBLIOGRAFÍA

- DOCAMPO, J.; COLODRON, V.: «Automatización de catálogos de material gráfico: la experiencia de la biblioteca Nacional de España» (sin publicar).
- AGENJO, X. (et al.): «La automatización de la CDU y los sistemas de gestión de bases de datos: el análisis de la estructura de la información». En: *FID 46th Conference and Congress: Madrid, 1992*.